

Claims

What is claimed is:

1. A computer-based method of computing a similarity function from a data set of objects, the method comprising the steps of:

- 5 obtaining a training set of objects;
- presenting a user with one or more subsets of objects based on the training set of objects, each subset comprising at least two objects of the data set;
- receiving user feedback regarding similarity between the one or more subsets of objects;
- 10 defining one or more sets of feature variables based on features in the one or more subsets of objects;
- defining one or more class variables in accordance with the user feedback; and
- constructing a similarity function which relates the one or more sets of feature variables to the one or more class variables.

15 2. The method of claim 1, wherein each of the sets of feature variables is defined by a pairwise subtraction of the features in the corresponding subset of objects.

 3. The method of claim 1, wherein the one or more class variables are based on one or more similarity values returned by the user.

20 4. The method of claim 1, wherein each of the sets of feature variables is defined as a parametric function of the features in the corresponding subset of objects.

 5. The method of claim 4, wherein one or more parameters of the function are determined in accordance with the user feedback.

6. The method of claim 1, wherein the similarity function is constructed for subsequent use in automatically determining the similarity between two or more objects of substantially unknown similarity.

7. Apparatus for computing a similarity function from a data set of objects, the apparatus comprising:

at least one processor operative to: (i) obtain a training set of objects; (ii) present a user with one or more subsets of objects based on the training set of objects, each subset comprising at least two objects of the data set; (iii) receive user feedback regarding similarity between the one or more subsets of objects; (iv) define one or more sets of feature variables based on features in the one or more subsets of objects; (v) define one or more class variables in accordance with the user feedback; and (vi) construct a similarity function which relates the one or more sets of feature variables to the one or more class variables; and

memory, operatively coupled to the at least one processor, for storing at least one of the data set of objects and the constructed similarity function.

8. The apparatus of claim 7, wherein each of the sets of feature variables is defined by a pairwise subtraction of the features in the corresponding subset of objects.

9. The apparatus of claim 7, wherein the one or more class variables are based on one or more similarity values returned by the user.

10. The apparatus of claim 7, wherein each of the sets of feature variables is defined as a parametric function of the features in the corresponding subset of objects.

11. The apparatus of claim 10, wherein one or more parameters of the function are determined in accordance with the user feedback.

12. The apparatus of claim 7, wherein the similarity function is constructed for subsequent use in automatically determining the similarity between two or more objects of substantially unknown similarity.

13. An article of manufacture for computing a similarity function from a data set of objects, the article comprising a machine readable medium containing one or more programs which when executed implement the steps of:

obtaining a training set of objects;

presenting a user with one or more subsets of objects based on the training set of objects, each subset comprising at least two objects of the data set;

receiving user feedback regarding similarity between the one or more subsets of objects;

defining one or more sets of feature variables based on features in the one or more subsets of objects;

defining one or more class variables in accordance with the user feedback; and

constructing a similarity function which relates the one or more sets of feature variables to the one or more class variables.

14. The article of claim 13, wherein each of the sets of feature variables is defined by a pairwise subtraction of the features in the corresponding subset of objects.

15. The article of claim 13, wherein the one or more class variables are based on one or more similarity values returned by the user.

16. The article of claim 13, wherein each of the sets of feature variables is defined as a parametric function of the features in the corresponding subset of objects.

